IMPORTANCE SAMPLING FOR TAIL PROBABILITIES

BY

SATISH IYENGAR

TECHNICAL REPORT NO. 440

FEBRUARY 6, 1991

PREPARED UNDER CONTRACT

N00014-89-J-1627 (NR-042-267)

FOR THE OFFICE OF NAVAL RESEARCH

DEPARTMENT OF STATISTICS

STANFORD UNIVERSITY

STANFORD, CALIFORNIA

IMPORTANCE SAMPLING FOR TAIL PROBABILITIES

BY

SATISH IYENGAR

TECHNICAL REPORT NO. 440
FEBRUARY 6, 1991

Prepared Under Contract
N00014-89-J-1627 (NR-042-267)
For the Office of Naval Research

Herbert Solomon, Project Director

DEPARTMENT OF STATISTICS
STANFORD UNIVERSITY
STANFORD, CALIFORNIA

1 Introduction

The assessment of statistical procedures often requires the evaluation of error probabilities that can be written as

$$\theta = \int_A f(x)dx\,, \tag{1}$$

where f is a probability density, and $A \subset R^p$ is a tail region typically of the form $\{x \in R^p : x_i \geq a_i\,, i = 1, \ldots, p\}$. When there is some special structure in such a problem, analytically tractable approximations or inequalities are available ([10],[15]); when this is not the case, however, Monte Carlo methods are often the only available option, especially for large p ([14]). Direct Monte Carlo uses the average of n independent replicates of $I(X \in A)$, where X is a random vector with density f and $I(E)$ is the indicator function of the event E. This unbiased estimate of θ has variance $(\theta - \theta^2)/n$, and squared coefficient of variation $(\theta - \theta^2)/n\theta$; thus, when θ is small, inordinately large n are needed to get a sufficiently accurate estimate. In such cases, importance sampling ([7],[11]) is a useful variance reduction technique which uses the expression

$$\theta = \int_A \frac{f(x)}{g(x)} g(x)dx\,, \tag{2}$$

for some "sampling density" g. This leads to another unbiased estimator which is the average of n independent replications of $\frac{f(Y)}{g(Y)} I(Y \in A)$ (where Y has density g) with variance $(\int_A \frac{f(x)^2}{g(x)} dx - \theta^2)/n$. The problem then is to find the sampling density which has a variance that is substantially smaller than that of direct Monte Carlo (that is, for which f/g is close to constant on A). In practice, generating observations from g and evaluating the ratio $f(x)/g(x)$ should also be easy; otherwise, any savings in variance reduction could be offset by the added cost of doing these calculations. This reduction in variance can be used in two ways: we can achieve a desired accuracy with a reduced sample size, or we can get a more accurate estimate with the same sample size. There is a large literature on importance sampling and other variance reduction techniques: for entries into this literature, see for example [1],[2],[3],[6], and [9]; for recent applications to engineering, see the references in [16].

It is well known ([7]) that the choice $g(x) = f(x)I(x \in A)/\theta$ gives an estimate with zero variance, but this observation is of limited use since it depends upon the unknown θ. An alternative approach is to restrict attention to a parametric family of functions g, and to choose an optimal one (or nearly so) from that family. In this paper, we study this technique for the evaluation of tail probabilities. Our interest is in the accurate and efficient evaluation of very small probabilities; thus, we seek methods with bounded coefficient of variation as the probability decreases (usually as the region of integration goes to infinity), and we assess their performance. Section 2 contains our notation and some preliminaries. Section 3 contains the univariate case and provides a heuristic for construction sampling density families. Section 4 contains the multivariate case, and provides interesting connections of this work with Mills' ratio, for which we provide an alternative definition. Section 5 contains proofs of selected results of this paper.

2 Notation and Preliminaries

Given an unbiased estimate $\hat{\theta}$ of θ, our criterion for assessing its performance is the squared coefficient of variation (cv^2):

$$cv^2(\hat{\theta}) = \frac{1}{n}\frac{var(\hat{\theta})}{\theta^2} = \frac{1}{n}\left(\frac{E(\hat{\theta}^2)}{\theta^2} - 1\right). \qquad (3)$$

This expression for cv^2 allows us to consider only the case $n = 1$, and study $E(\hat{\theta}^2)$ as a function of the set A. Let $N_p(0, \Sigma)$ denote a normal distribution with mean 0 and covariance $\Sigma = (\rho_{ij})$ with $\rho_{ii} = 1$; when Σ is nonsingular, $\phi_p(x; \Sigma)$ is the corresponding density. When $p = 1$, let $\Phi(x)$ and ϕ denote the standard normal distribution and density, respectively. A $N_p(0, I)$ variate, where I stands for the identity matrix, will be denoted Z; its dimension will be clear from the context.

A function which appears in many calculations below is Mills' ratio ([4]):

$$M(x) = \frac{\Phi(-x)}{\phi(x)} \tag{4}$$

and its multivariate version ([12])

$$M(x; \Sigma) = \frac{P(X \geq x)}{\phi_p(x; \Sigma)}, \tag{5}$$

where X is a $N_p(0, \Sigma)$ variate, and $x \geq y$ means $x_i \geq y_i$ for all i. The properties of these functions that we need are given in the appendix. In Section 4, we introduce another definition of a multivariate normal Mills' ratio and study its properties there.

3 The Univariate Case

Even though importance sampling is much more important for high-dimensional probability integrals, we begin with a treatment of the univariate case for several reasons. First, it allows for a careful analytical examination of the degree of improvement that importance sampling affords. Next, it provides guidelines for choosing families of sampling density for the multivariate case. Finally, our discussion here provides justification of some of the results in [16].

We start with normal tail probabilities: let $\theta = \Phi(-a)$ for large positive a, for which direct Monte Carlo has $cv^2 = \Phi(-a)^{-1} - 1$. Two families of sampling density that have been suggested in this case ([16]) are $\{\phi(x-\mu) : \mu \in R\}$ and the $\{\frac{1}{\sigma}\phi(\frac{x-\mu}{\sigma}) : \mu \in R,\ \sigma \geq 0\}$. For the first family, after a change of variables we have

$$\theta = e^{-\mu^2/2} \int_{a-\mu}^{\infty} e^{-\mu x} \phi(x) dx\,, \tag{6}$$

suggesting the estimator

$$\hat{\theta}_\mu = e^{-\mu^2/2} e^{-\mu Z} I(Z \geq a - \mu)\,, \tag{7}$$

which has

$$E(\hat{\theta}_\mu^2) = e^{\mu^2} \Phi(-\mu - a)\,. \tag{8}$$

A special case is $\mu = a$: since $M(x) \sim 1/x$ for large x, we have

$$cv^2(\hat{\theta}_a) \sim \sqrt{\frac{\pi}{2}} a - 1 \tag{9}$$

It turns out (proofs of (10 to (14) are in the appendix) that the optimal members of each of the two sampling density families do not provide much more of an improvement as $a \to \infty$. For the first family, the optimal value, μ^*, of μ satisfies $2\mu M(a + \mu) = 1$, and since $M(x) \geq \frac{x}{1+x^2}$ for $x > 0$, $a < \mu^* < a + \sqrt{1 + a^2}$, so that

$$\frac{var(\hat{\theta}_{\mu^*})}{var(\hat{\theta}_a)} \to 1 \text{ as } a \to \infty. \tag{10}$$

For the second family, the optimal values are $\mu^* = (a + \sqrt{a^2 + 2})/2$ and $\sigma^* = 1/\sqrt{2}$, so that

$$\frac{var(\hat{\theta}_{\mu^*,\sigma^*})}{var(\hat{\theta}_a)} \to \frac{1}{\sqrt{2}} \text{ as } a \to \infty. \tag{11}$$

Thus, the optimal members of these sampling density families provide a moderate improvement over $\phi(x-a)$, which provides a substantial improvement over direct Monte Carlo.

Notice, however, that the optimal estimators above (and $\hat{\theta}_a$) each have coefficient of variation increasing at the rate $\sqrt{a}$ as a increases, so that the relative error increases without bound as the threshold increases. In [16], the exponential density is shown to overcome this problem. Writing

$$\theta = \frac{\phi(a)}{a}\int_0^\infty e^{-x^2/2a}e^{-x}dx\,, \tag{12}$$

yields the estimator

$$\hat{\theta} = \frac{\phi(a)}{a}e^{-X^2/2a}\,, \tag{13}$$

where X has the exponential density e^{-x} for $x \geq 0$. We have

$$cv^2(\hat{\theta}) = \frac{M(a/\sqrt{2})}{M(a)^2 a\sqrt{2}} - 1 \sim \frac{2}{a^2} \text{ as } a \to \infty\,, \tag{14}$$

so that the relative error actually decreases with the tail probability.

More generally, a heuristic method for choosing a sampling density (based on a "simulacrum") is provided in [16] for such problems. Here, we propose an alternative heuristic based on some elementary calculations. If X has density f, l'Hôpital's rule says that with suitable regularity, the asymptotic behavior of $P(X>a)/f(a)$ is the same as that of $-f(a)/f'(a)$. Using a tractable approximation $r(a)$ for $-f(a)/f'(a)$, we get

$$\theta = P(X>a) = r(a)f(a)\int_a^\infty \frac{f(x)}{r(a)f(a)}dx = r(a)f(a)\int_0^\infty \frac{f(x+a)}{r(a)f(a)}dt. \tag{15}$$

Under the same regularity conditions, the last integral in (15) approaches 1 as $a \to \infty$; thus, it is bounded away from 0, and estimating it with good *relative* accuracy can be done using importance sampling. In fact, it is a rather remarkable fact that the phenomenon observed in (14) is quite general: for a wide class of problems, the coefficient of variation not only remains bounded, but it goes to zero , and hence the relative accuracy improves as the threshold a increases. In addition, this method is feasible since the calculation of $r(a)$ depends on the differentiation of the density rather than its integration; since the tail behavior F is already captured by $r(a)f(a)$, the evaluation of the remaining integral by Monte Carlo just provides a correction term.

In practice, we use (15) or either of the following two expressions for θ:

$$\theta = r(a)f(a)\int_0^\infty \frac{f(\frac{x}{a}+a)}{ar(a)f(a)}dx = r(a)f(a)\int_0^\infty \frac{af(a(x+1))}{r(a)f(a)}dx\,, \tag{16}$$

and use as a sampling density one that matches the tail behavior of the integrand. The calculation for the normal in (12-14) uses the first expression in (16). Instead of giving a general result, we present some examples to illustrate the calculations (in all examples, cv^2 tends to 0 as a tends to ∞).

Example 1: Let $f_\beta(x) = x^\beta e^{-x}/\Gamma(\beta)$ so that $r(a) = \frac{a}{a-\beta} \sim 1$, and using (15) we have

$$\theta = f_\beta(a)\int_0^\infty (\frac{x}{a}+1)^\beta e^{-x}dx \tag{17}$$

yielding the estimator

$$\hat{\theta} = f_\beta(a)(\frac{X}{a}+1)^\beta\,, \tag{18}$$

where X has a standard exponential distribution. Notice that the tail of "generalized Gaussian distribution" ([16]) with density proportional to $exp(-x^{-\beta}/\beta)$ can be reduced to this case by a change of variables. Notice also that when $\beta = 0$, the variance is identically 0.

Example 2: For the t distribution with k degrees of freedom, the density $f_k(x)$ is proportional to $(1+\frac{x^2}{k})^{-(k+1)/2}$, so that using the second equation in (16) we have

$$\hat{\theta} = \frac{a}{k} f_k(a) [\frac{(k+a^2)X^2}{(k+a^2X^2)}]^{\frac{k+1}{2}}\,, \tag{19}$$

where X has density k/x^{k+1} for $x \geq 1$. A close examination of this example with $k=1$ (Cauchy distribution) shows that this method works even when l'Hôpital's rule does not give the right constant in the asymptotics: in this case the two differ by a factor of 2, $r(a)f(a) \sim 1/2\pi a$, while $P(X > a)/f(a) \sim 1/\pi a$.

The use of (6) with $\mu = a$ is an instance of what is called "improved importance sampling" (IIS) [16]; we now show that the method based on l'Hôpital's rule described above is better than IIS. Let $T \geq 0$ be a random variable with density f, which is decreasing; then IIS writes

$$\theta = \int_a^\infty f(x)dx = \int_0^\infty \frac{f(x+a)}{f(x)} f(x)dx \tag{20}$$

to suggest the estimator $\hat{\theta}_a = f(T+a)/f(T)$. Now suppose that $f(x+a)/f(x) \leq h(a)$ for some function $h \leq 1$ for all $x \geq 0$. When f is normal, $h(a) = e^{-a^2/2}$; when f is logistic, $h(a) = e^{-a}$; and when f is a t density with k degrees of freedom, $h(a) = 1$. Since

$$E(\hat{\theta}_a^2) = \int_0^\infty \frac{f(x+a)}{f(x)} f(x+a)dx \leq \int_a^\infty h(a)f(x)dx = h(a)\theta\,, \tag{21}$$

we have

$$\frac{E(\hat{\theta}_a^2)}{\theta^2} \leq \frac{h(a)}{\theta} \quad \text{and} \quad \frac{var(\hat{\theta}_a)}{var(\hat{\theta}_0)} \leq h(a)\,, \tag{22}$$

where $\hat{\theta}_0$ is the direct Monte Carlo estimate. Thus, even though IIS is an improvement over direct Monte Carlo for such problems, (22) shows that unless (as $a \to \infty$) h goes to 0 faster than θ does, the IIS estimator's coefficient of variation will tend to infinity.

4 The Multivariate Case

We now turn to the multivariate case, concentrating on the multivariate normal. For this discussion, we principally deal with the spherically symmetric case, $\Sigma = I$. There is no loss of generality because of the following simple transformation: if $X \in R^p$ has covariance matrix Σ, then $P(X \in A) = P(\Sigma^{-\frac{1}{2}}X \in \Sigma^{-\frac{1}{2}}A)$, and $\Sigma^{-\frac{1}{2}}X$ is spherically symmetric; also, a simulation which generates multivariate observations usually starts by generating the spherically symmetric variates.

Now suppose that Z is a $N_p(0, I)$ variate and $A \in R^p$ is a closed convex set (such as an orthant, a rectangle, or a sphere). Suppose that A does not contain the origin; then there is a point $a \in A$ that is closest (in ordinary Euclidean distance) to the origin: $| a | \leq | x |$ for all $x \in A$. As before, let $\theta = P(X \in A)$. Analogous to (6) we have the estimators

$$\hat{\theta}_0 = I(Z \in A) \tag{23}$$

and

$$\hat{\theta}_a = e^{-|a|/2} e^{-a'Z} I(Z \in A - a), \tag{24}$$

where $A - a = \{x - a : x \in A\}$.

We first show that $\hat{\theta}_a$ is better than $\hat{\theta}_0$. Since A is convex and does not contain 0, A is contained in the half space $\{x : a'x \geq a'a\}$. Thus,

$$E(\hat{\theta}_a^2) = e^{a'a} P(Z \in A + a) = e^{a'a/2} \int_A e^{-a'x} \phi_p(x; I)\, dx \leq e^{-a'a/2}\theta = e^{-a'a/2} E(\hat{\theta}_0^2). \tag{25}$$

Thus, as $| a | \to \infty$, $\hat{\theta}_a$ provides a vast improvement over $\hat{\theta}_0$. The convexity of A is crucial, as seen by the following counterexample. Let $A = \{x \in R^2 : | x | \geq r\}$, and let $\theta = P(Z \in A)$, so that the direct Monte Carlo estimate has second moment θ. If we use the sampling density $\phi_p(x-a;I)$, where a is any point on the circle $\{x \in R^2 : |x| = r\}$, a routine calculation shows that the new estimator has second moment $e^{r^2/2}P(\chi^2(1) \geq r^2)$, which is at least $e^{r^2/2}P(\chi^2 \geq r^2 = e^{r^2/2}\theta$, since the non-central chi-square is stochastically larger than the central one ($\chi^2(\psi^2)$ denotes a non-central chi-square variate with non-centrality parameter ψ^2).

It can be shown that analogous to (14), $cv^2(\hat{\theta}_a)$ increases at a rate proportional to $|a|^p$, and so we generalize the approach leading to (15) to get estimators with bounded cv^2 after some preparation. We introduce here a definition of a multivariate Mills' ratio, which arises in our work below. Let A and a be as before, and let

$$M(A;I) = P(Z \in A)/\phi_p(a;I). \tag{26}$$

Notice that we have set $\Sigma = I$ in (26); alternatively, we can define

$$M(A;\Sigma) = P(X \in A)/\phi_p(a;\Sigma). \tag{27}$$

for X a $N_p(0,\Sigma)$ variate, where $a \in A$ minimizes the Mahalanobis distance from the origin. That is,

$$a'\Sigma^{-1}a \leq x'\Sigma^{-1}x \tag{28}$$

for all $x \in A$. For a given A, finding a is a quadratic programming problem, and can be solved by standard iterative methods [5]. Now consider the case in which A is the orthant $\{x \in R^p : x \geq b\}$ for some $b \in R^p$. For evaluating $P(X \in A)$ where X is a $N_p(0,\Sigma)$ variate, the sampling density $\phi_p(x-b;I)$ has been suggested ([2]). In addition, the multivariate generalization of Mills' ratio proposed in [12] is $P(X \in A)/\phi_p(x-b;\Sigma)$, and approximations to it have been studied in [8],[13]. We now propose new alternatives to each of these.

First, instead of using $\phi_p(x-b;\Sigma)$, (25) suggests the use of the sampling density $\phi_p(x-a;\Sigma)$, where a is given by (28). A simple example clarifies the situation. Suppose that $p=2$, $A=\{x : x_1 \geq a_1, x_2 \geq 0\}$. If $corr(X_1,X_2)=\rho$, then $a=(a_1,a_1\rho)$ while $b=(a_1,0)$. By (25), the use of a is better than direct Monte Carlo; furthermore, it can be shown here that the use of a is better than that of b, and that the use of b can actually be worse than direct Monte Carlo for sufficiently large ρ.

More generally, for any convex A not containing the origin, for $\Sigma = I$ we can rotate the axes to get $a = \mid a \mid e_1$, where e_1 is the unit vector in the x_1 direction, so that A is contained in the half space $\{x : x_1 \geq \mid a \mid\}$.

$$P(Z \in A) = \phi_p(a;I) \int_{A-a} e^{-a'x - x'x/2} dx \tag{29}$$

or after some reductions,

$$P(Z \in A) = \frac{\phi_p(a;I)}{\mid a \mid} \int_{T_a(A-a)} e^{-t_1^2/2|a|} exp(-t_1 - y'y/2) dt_1 dy \tag{30}$$

where $y=(t_2,\ldots,t_p)$ and T_a is the matrix $diag(\mid a \mid, 1, \ldots, 1)$.

The expression (30) yields several results. First, it shows that the sampling density should be proportional to $exp(-t_1 - y'y/2)$; using methods similar to those leading to (14), it can be shown that the coefficient of variation decreases to zero if $A-a$ is a cone, or if $T_a(A-a) = A-a$. Next, it provides an inequality for our Mills' ratio.

Proposition 1 : *For convex A,*

$$M(A;I) < \frac{e^{-|a|^2}/2}{\sqrt{2\pi} \mid a \mid}. \tag{31}$$

An asymptotic expansion analogous to that of the one-dimensional case can be derived also by expanding the integrand $e^{-t_1^2/2|a|}$ in (30) and integrating term by term.

5 Appendix

Proof of (10): Let $g(\mu)$ be $E(\hat{\theta}_\mu^2)$ from (8); to find the optimal value μ^* of μ, we solve $g'(\mu) = 0$ or

$$2\mu M(\mu + a) = 1. \tag{32}$$

Clearly, $\mu^* \geq 0$; next, if $0 \leq \mu \leq a$ then

$$M(\mu + a) - \frac{1}{2\mu} < \frac{1}{\mu + a} - \frac{1}{2a} \leq 0\,, \tag{33}$$

so that $\mu^* > a$. Next, since for $x > 0$ we have $M(x) > \frac{x}{1+x^2}$,

$$\frac{1}{2\mu^*} = M(a + \mu^*) > \frac{a + \mu^*}{1 + (a + \mu^*)^2}\,, \tag{34}$$

we have $\mu^* < \sqrt{a^2 + 1} < a + \frac{1}{2a}$. To show the asymptotic equivalence of $\hat{\theta}_{\mu^*}$ and $\hat{\theta}_a$, note that we only need consider second moments instead of variances. Using (32), to get the equation below, and writing μ for μ^*, we have

$$1 \geq \frac{E(\hat{\theta}_\mu)}{\hat{\theta}_a} = \frac{e^{\mu^2}\phi(a+\mu)}{2\mu e^{a^2}\Phi(-2a)} \sim \frac{a}{\mu} exp(\frac{1}{2}(\mu - a)^2)\,, \tag{35}$$

which tends to 1 as $a \to \infty$.

Proof of (11): Writing $\tau = 1/\sigma$, we have

$$\theta = \int_{\tau(a-\mu)}^{\infty} \frac{1}{\tau} \frac{\phi(\frac{x}{\tau} + \mu)}{\phi(x)} \phi(x) dx. \tag{36}$$

For $\hat{\theta}_{\tau,\mu}$ derived from this expression,

$$E(\hat{\theta}_{\tau,\mu}^2) = \frac{1}{\tau\sqrt{2 - \tau^2}} exp(\frac{\mu^2\tau^2}{2 - \tau^2})\Phi(-\frac{(2 - \tau^2)a + \mu\tau^2}{\sqrt{2 - \tau^2}}). \tag{37}$$

Using the same methods as in the proof of (10), it can be shown that the optimal values of μ and τ satisfy (i) $a < \mu^* < a + \frac{1}{2a}$, and $1 \leq \tau^* < 2$. Writing μ and τ for μ^* and τ^*, and using the approximation $M(x) \sim 1/x$, it is easy to show that

$$\frac{E(\hat{\theta}^2_{\tau,\mu})}{E(\hat{\theta}^2_a)} \sim \frac{1}{\tau} exp(\frac{\tau^2}{2}(\mu - a)^2). \tag{38}$$

Since $\mu - a \leq \frac{1}{2a}$ and $\tau \geq 1$, the ratio in (38) is bounded between 1/2 and 1 as $a \to \infty$, so that the optimal member of this family is asymptotically equivalent (up to a constant) to $\hat{\theta}_a$. In fact, more involved calculations show that the optimal values are those given above in (11).

References

[1] Evans, M. (1988) Monte Carlo Computation of Marginal Posterior Distributions. University of Toronto Statistics Department Tech. Report 11.

[2] Evans, M. and Swartz, T. (1988) Monte Carlo Computation of Some Multivariate Normal Probabilities. *J. Statist. Comp. Simul.* **30**, 117-128.

[3] Evans, M. and Swartz, T. (1988) Sampling from Gauss Rules. *SIAM J. Scient. Statist. Comp.* **9**, 950-961.

[4] Feller, W. (1968) *An Introduction to Probability Theory and its Applications, vol. 1.* Wiley. New York.

[5] Fletcher, R. (1981) *Practical Methods of Optimization, vol. 2: Constrained Optimization.* Wiley. Chichester.

[6] Geweke, J. (1988) Acceleration Methods for Monte Carlo Integration in Bayesian Inference. *Computer Science and Statistics: Proc. of 20th Symposium on the Interface.* 587-592.

[7] Hammersley, I. and Handscomb, D. (1964) *Monte Carlo Methods.* Chapman and Hall. London.

[8] Iyengar, S. (1986) On a Lower Bound for the Multivariate Normal Mills' Ratio. *Ann. Probab.* **14**, 1399-1403.

[9] Kloek, T. and van Dijk, H. (1978) Bayesian Estimates of Equation System Parameters: An Application of Integration by Monte Carlo. *Econometrica.* **46**, 1-20.

[10] Martynov, G. (1980) Evaluation of the Normal Distribution Function. *J. Soviet Math.* **17**, 1857-1875.

[11] Rubinstein, R. (1981) *Simulation and the Monte Carlo Method.* Wiley, New York.

[12] Savage, I.R. (1962) Mills' Ratio for Multivariate Normal Distributions. *J. Res. Nat. Bur. Siand.* **B66**, 93-96.

[13] Steck, G. (1979) Lower Bounds for the Multivariate Normal Mills' Ratio. *Ann. Probab.* **7**, 547-551.

[14] Thisted R. (1988) *Elements of Statistical Computing.* Chapman and Hall. New York.

[15] Tong,Y.L. (1980) *Probability Inequalities in Multivariate Distributions.* Academic Press. New York.

[16] Wessell, A., Hall, E. and Wise, G. (1988) Some Comments on Importance Sampling. *Proceedings of the 22nd Conference on Information Sciences and Systems.*

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER 440	2. GOVT ACCESSION NO.	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) Importance Sampling for Tail Probabilities		5. TYPE OF REPORT & PERIOD COVERED TECHNICAL REPORT
		6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s) Satish Iyengar		8. CONTRACT OR GRANT NUMBER(s) N00014-89-J-1627
9. PERFORMING ORGANIZATION NAME AND ADDRESS Department of Statistics Stanford University Stanford, CA 94305		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS NR-042-267
11. CONTROLLING OFFICE NAME AND ADDRESS Office of Naval Research Statistics & Probability Program Code 1111		12. REPORT DATE February 6, 1991
		13. NUMBER OF PAGES 16
14. MONITORING AGENCY NAME & ADDRESS(if different from Controlling Office)		15. SECURITY CLASS. (of this report) UNCLASSIFIED
		15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

APPROVED FOR PUBLIC RELEASE: DISTRIBUTION UNLIMITED.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Importance sampling, Tail probabilities, Mills' ratio

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

Our interest is in the accurate and efficient evaluation of very small probabilities. Thus, we seek methods with bounded coefficient of variation as the probability decreases and we assess their performance. The univariate case is explored for a heuristic for construction of sampling density families. The multivariate case provides interesting connections with Mills' ratio.

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE

S/N 0102-014-6601

UNCLASSIFIED